Supporting Online Material for

Accessible Reproducible Research

Jill P. Mesirov[1]

[1]Broad Institute of Massachusetts Institute of Technology and Harvard University,

7 Cambridge Center, Cambridge MA 02142, USA.

1. Brief technical description of the GenePattern-Word RRS
2. Figs. S1-S4

1. Brief technical description of the GenePattern-Word RRS:

The GenePattern-Word RRS was possible because of a convergence of technologies: specifically, the GenePattern Web services architecture, the Microsoft .NET Web services architecture and the Microsoft Office 2007 suite support for Word 2007 add-ins and Open XML Document format. These allowed creation of a GenePattern .NET client that communicates with the GenePattern server via the Simple Object Access Protocol (SOAP), seamless integration of this client into the Word 2007 environment, and storage in the Open XML formatted document of the relationships between document figures or text and GenePattern pipelines. The Open XML document can contain the pipeline definition, data files, and analysis and visualization module executables for replication of the analysis on any GenePattern server, or just a reference to the pipeline on a particular server.

The work to develop the GenePattern-Word RRS did not require any changes to the GenePattern software and a similar approach could be used for any scientific software package that can 'speak' to a Web services interface. The critical factors in this experiment were the openness of the GenePattern and Word software interfaces and the ability of GenePattern to capture and package a scientist's session, including the tools, settings, and data that were used. The GenePattern-Word RRS is an open source add-in to the Microsoft Office application, freely available at http://genepatternwordaddin.codeplex.com.

2. Figures S1-S4. The GenePattern Reproducible Research Document:

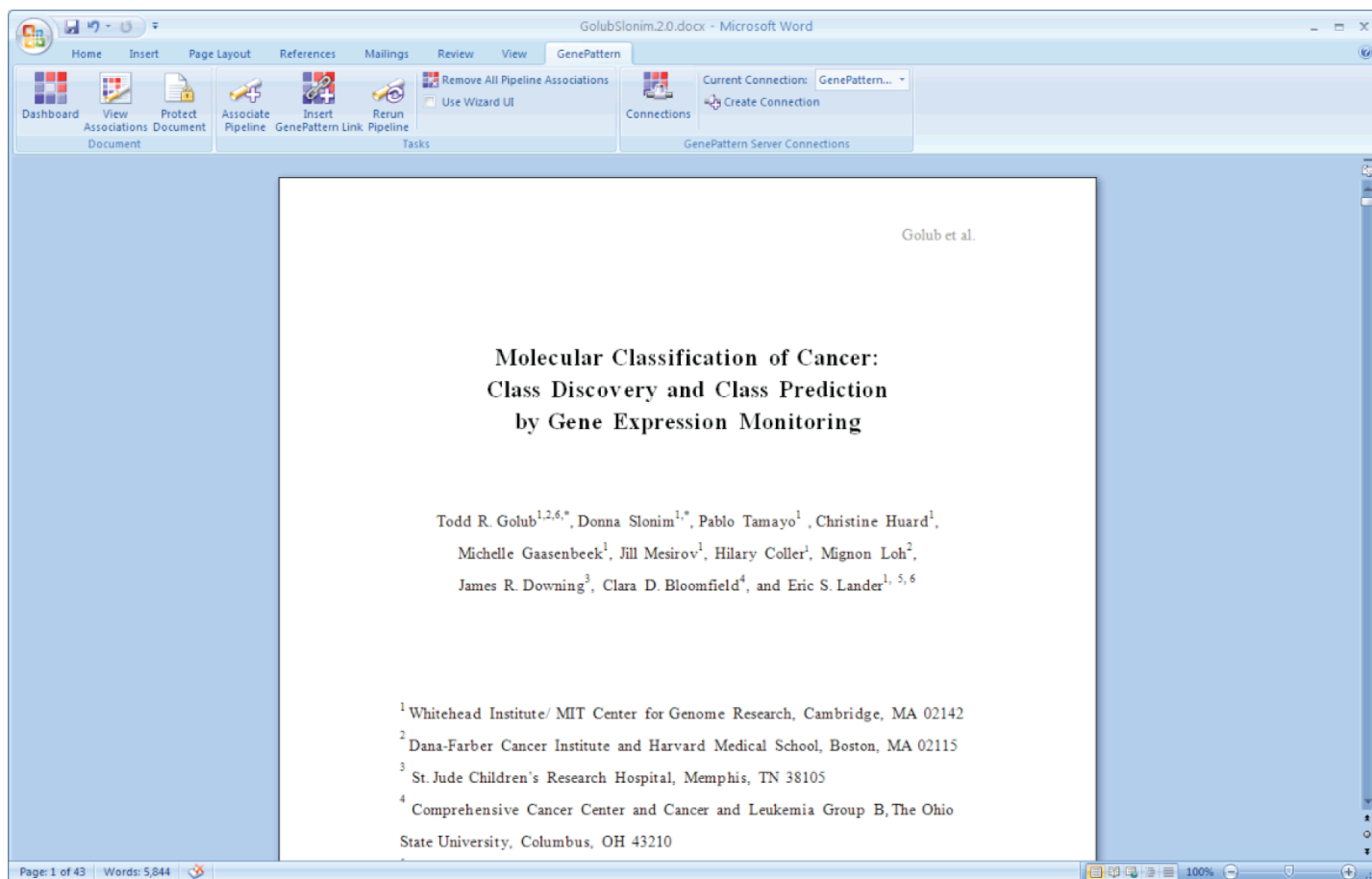Four screen shots of a Word document containing GenePattern pipeline associations.

Fig S1. A Word document with an active **GenePattern ribbon** that gives the user access to GenePattern functionality within the Word application.
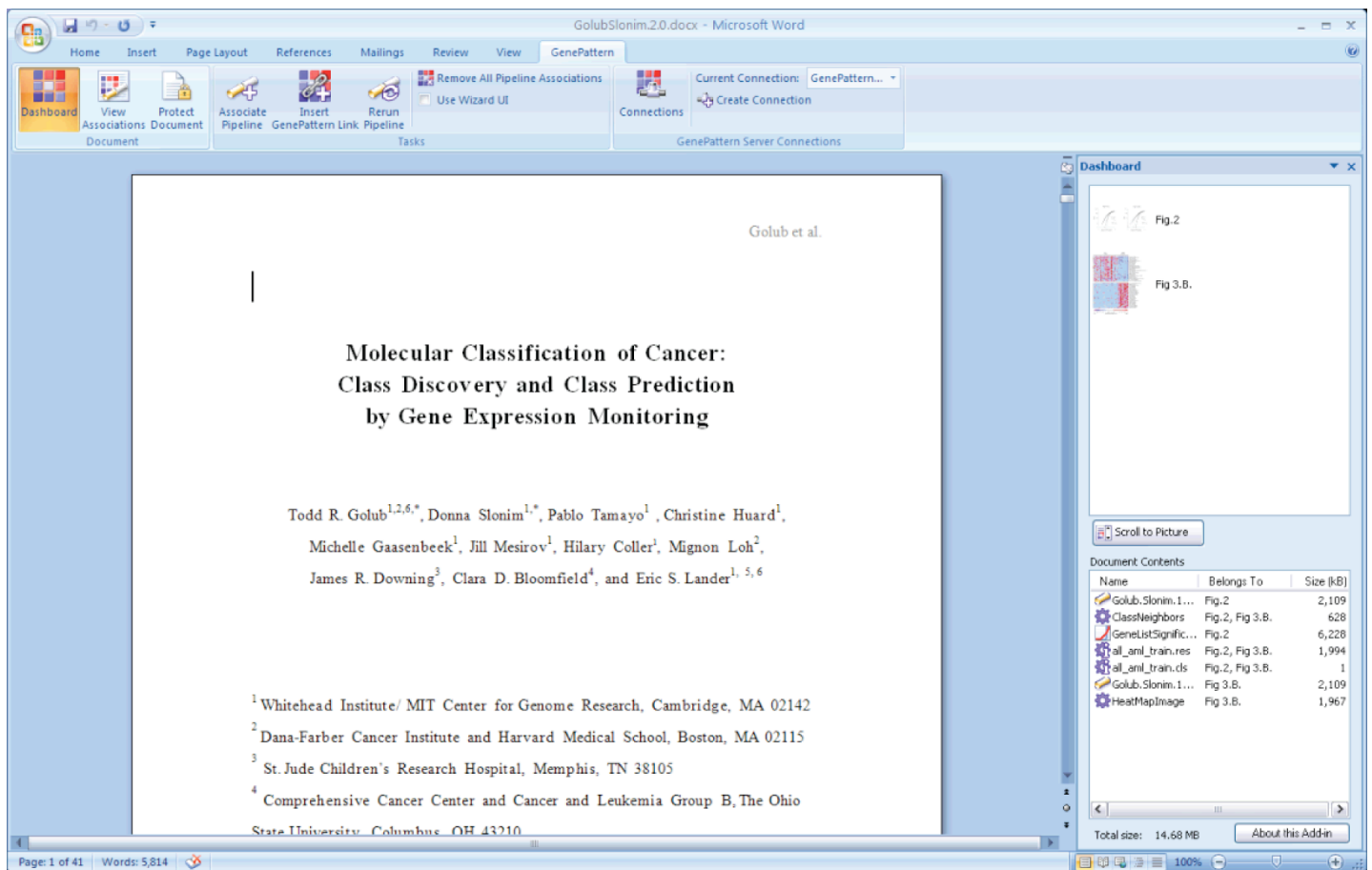
Fig S2. A reader has opened a task pane, visible on the right, that contains the GenePattern **dashboard.** The top section displays icons of figures within the paper that have associated pipelines. The bottom section lists the entire GenePattern content of the paper. This includes pipelines and their constituent analytic and visualization modules. Each of these different forms of content has a specific icon representation and a name.

Fig S3. A reader has scrolled down the document and selected a figure with a pipeline association. On the right is the **associated pipelines** task pane. The top section lists all the pipelines associated with that figure (there can be more than one). The middle section lists all the steps of the pipeline that the user highlighted in the upper part of the pane. There is also a tab for viewing any author-provided annotation. The bottom of the pane has three tabs. The visible tab lists all the output files for the pipeline. The other two list the pipeline parameters and output images.

Fig S4. This screen shot shows the results of rerunning an embedded pipeline and inserting the image back into the document as an appendix. The entire rerun was accomplished without ever leaving the Word application or the document itself. The appendix includes a user-provided description of the rerun including the pipeline association name, date and time of the run, and any other comments. When results are saved to the appendix all user-configured parameter information, including any input files, is recorded. This particular rerun did not have any user configurable parameters. Both input and result files are embedded or identified by URLs.